Guidelines for creating artificial neural network empirical interatomic potential from first-principles molecular dynamics data under specific conditions and its application to α-Ag$_2$Se

Kohei Shimamura (iD), Shogo Fukushima, Akihide Koura, Fuyuki Shimojo, Masaaki Misawa, Rajiv K. Kalia, Aiichiro Nakano (iD), Priya Vashishta, Takashi Matsubara, and Shigenori Tanaka

View Online   Export Citation   CrossMark

AIP Publishing

Guidelines for creating artificial neural network empirical interatomic potential from first-principles molecular dynamics data under specific conditions and its application to α-Ag$_2$Se

View Online   Export Citation   CrossMark

Kohei Shimamura,[1,a] Shogo Fukushima,[2] Akihide Koura,[2] Fuyuki Shimojo,[2] Masaaki Misawa,[3] Rajiv K. Kalia,[4] Aiichiro Nakano,[4] Priya Vashishta,[4] Takashi Matsubara,[1] and Shigenori Tanaka[1]

AFFILIATIONS

[1] Graduate School of System Informatics, Kobe University, Kobe 657-8501, Japan

[2] Department of Physics, Kumamoto University, Kumamoto 860-8555, Japan

[3] Faculty of Science and Engineering, Kyushu Sangyo University, Fukuoka 813-8503, Japan

[4] Collaboratory for Advanced Computing and Simulations, University of Southern California, Los Angeles, California 90089-0242, USA

[a] Author to whom correspondence should be addressed: shimamura@port.kobe-u.ac.jp

ABSTRACT

First-principles molecular dynamics (FPMD) simulations are highly accurate, but due to their high calculation cost, the computational scale is often limited to hundreds of atoms and few picoseconds under specific temperature and pressure conditions. We present here the guidelines for creating artificial neural network empirical interatomic potential (ANN potential) trained with such a limited FPMD data, which can perform long time scale MD simulations at least under the same conditions. The FPMD data for training are prepared on the basis of the convergence of radial distribution function [$g(r)$]. While training the ANN using total energy and atomic forces of the FPMD data, the error of pressure is also monitored and minimized. To create further robust potential, we add a small amount of FPMD data to reproduce the interaction between two atoms that are close to each other. ANN potentials for α-Ag$_2$Se were created as an application example, and it has been confirmed that not only $g(r)$ and mean square displacements but also the specific heat requiring a long time scale simulation matched the FPMD and the experimental values. In addition, the MD simulation using the ANN potential achieved over 10^4 acceleration over the FPMD one. The guidelines proposed here mitigate the creation difficulty of the ANN potential, and a lot of FPMD data sleeping on the hard disk after the research may be put on the front stage again.

I. INTRODUCTION

The large-scale and long time scale first-principles molecular dynamics (FPMD) simulations are not practical, considering their large computational complexity [more than $O(N^3)$ for the number of atoms N]. For the system consisting of several hundred atoms, several picosecond simulations would be standard, but it is difficult to get enough statistics on this scale. In particular, in order to obtain the convergence of physical quantities such as specific heat and thermal conductivity calculated from energy fluctuations, it is necessary to perform the MD simulation for quite a long time.[1] Therefore, it is the mainstream to employ the classical MD simulation instead of the FPMD method, where an FP potential is replaced by an empirical interatomic potential with low computational complexity with sacrificing precision. Under such circumstances, an artificial neural network (ANN) has been proposed as a promising method for

resolving this issue. It uses the universal approximation of the ANN to learn FPMD simulations as training data to construct an empirical potential (referred to as an ANN potential).[2] Since the computational cost of MD simulations using the ANN potential (referred to as ANN-MD) is basically $O(N)$, a high-speed MD simulation is possible while retaining the accuracy of the FPMD one. Applications to MD have been not only to covalent crystal bulks[3] and ion crystal bulk/cluster[4,5] but also recently to systems with a high degree of freedom, such as water,[6,7] aqueous electrolyte solutions,[8] and solid-liquid interfaces,[9] where proton transfer between molecules, which was difficult with the conventional empirical potentials, is expressed with the FPMD precision. The ability of ANN to imitate the FP accuracy can be exploited in many research fields, for example, methods and program packages for rapidly predicting the properties of a wide range of molecules that can be applied to chemical structure search and drug screening have been actively developed.[10,11] As understood from these active research studies, the ANN (or ANN potential) figures prominently in the physical chemistry field.

Furthermore, recently, the development of interatomic potential with high transferability by active learning technology is also progressing.[2,12–14] Using the active learning, it might be possible to achieve efficient training of the ANN potential by quantifying uncertainty of some important physical quantities from the output of the potential (for example, estimation error of total energy[2] or atomic force[14]) and selecting data and algorithm to minimize the value.

Meanwhile, what we would like to mention here is what we should keep in mind for creating the ANN potential when we have FPMD data with specific conditions (e.g., hundreds of atoms and several picoseconds). Note that this study deals with robust potentials, at least under the specific conditions. Due to the large computational complexity, the studies conducted using FPMD simulations would be basically limited to specific temperature, pressure, and number of atoms as well as short time scale. There may be cases that one is interested only in specific conditions (e.g., temperature, pressure, and density) tailored to the experimental conditions or the standard condition if one assumes *in vivo* circumstances. In such cases, one often wants to investigate the atomic dynamics beyond the time domain that FPMD simulations can reach. An example is the calculation of the above-mentioned physical quantities such as the specific heat and thermal conductivity which need a long time to converge. Therefore, if ANN potentials can be created from the limited data, it will be possible to obtain more findings than those from only FPMD results. In addition, accumulation of such know-how is important because FPMD data being saved after the studies can be utilized again. We also recognize that the active learning above[14] is effective to create transferable potentials. However, in this method, searching the state space comprehensively by changing the temperature and the structure variously is necessary. Unless one is familiar with machine learning, it is difficult to judge whether or not the scheduling of exhaustive search is sufficient. On the other hand, if a robust ANN potential is created under the same specific conditions as the training data, it can be used for the starting point (template) with the active learning. Furthermore, when performing ANN-MD simulations under *different* conditions from the training data, with uncertainty defined by active learning, it seems possible to investigate whether physicochemical information that the training data includes can access the various conditions. Such an analysis method may be useful to understand the systems with phase transitions. (The

prospect of application to structural phase transition is described in Sec. VI of the supplementary material.)

Here are the four empirical guidelines that we consider important:

1. Prepare FPMD data as training data (including total energy, atomic force, pressure, and atomic Cartesian coordinates) so that the radial distribution functions [$g(r)$] are converged.
2. Not only total energy but also atomic force should be reproduced while training the ANN.
3. The error of pressure must always be monitored during training.
4. Prepare additional FPMD data to correct interaction between two atoms approaching each other.

The issue 1 is a criterion to determine whether training data are sufficient (e.g., we prepared 1000-step FPMD data). If $g(r)$ is appropriately acquired, it indicates that detailed information on the local structure determined by the interaction between atoms is obtained. It would be thus physically relevant to judge that the FPMD data with $g(r)$ being converged are sufficient as training data.

Next is concerned with the issue 2. Recently, there are an increasing number of studies that have proposed ANN potentials trained with not only total energy but also atomic force. We also consider that training with atomic force is essential to achieve the training with precision of both global (i.e., total energy) and local aspects. However, more important thing is the issue 3. As described later, even if the prediction of total energy and atomic force shows good accuracy, the prediction of pressure is often poor. The predicted pressure does not match that of the FPMD data if both global and local precisions of the ANN potential are not sufficient.[15] That is why monitoring the pressure error in training must be essential.

However, there is still a problem when considering the pressure error, which is a failure that occurs in ANN-MD simulations due to the insufficient reprehensibility of interatomic repulsion. We overcome it by adding a small amount of FPMD data to correct the interaction between atoms that are close to each other. Although the additional FPMD data may be needed in this way, one can create a robust ANN potential capable of performing long time scale ANN-MD simulations (e.g., nanosecond order) under the same conditions as FPMD data.

These four guidelines are used for the improvement of the ANN potential creation code, the Atomic energy network (Aenet) developed by Artrith and Urban.[16] In addition, we use our own code QXMD[17,18] to perform FPMD and ANN-MD simulations.

This paper attempts to create ANN potentials for a superionic conductor, α phase of Ag_2Se (α-Ag_2Se) according to the guidelines above. Ag_2Se is a chalcogenide recently attracting attention as a thermoelectric material.[19–21] An empirical interatomic potential other than the ANN potential has already been proposed for α-Ag_2Se, which shows good accuracy with respect to the diffusion coefficient of Ag and the $g(r)$ around Se.[22] However, the $g(r)$ between Ag-Ag largely deviates from the FPMD simulation and experiment.[22] This discrepancy is a hindrance to clarify the atomic mechanism of the spike of specific heat and thermal conductivity reported near the phase transition temperature (T_c = 406 K) between low temperature β phases (β-Ag_2Se).[20,23,24] As described above, since a long time

scale MD simulation is required to calculate the specific heat and thermal conductivity, we create a robust ANN potential that can reproduce the structure of Ag correctly and calculate the specific heat with respect to α-Ag$_2$Se.

II. METHOD OF CALCULATION

A. Artificial neural network (ANN)

Here, a general method of creating the ANN potential is described. The feedforward neural network (FFNN) is the simplest type of ANN,[25,26] where only forward propagation of information from input nodes to output nodes is allowed. It has been applied to MD simulations for more than a dozen years.[27–29] Figure 1 shows the structure of a simple FFNN comprising an input layer, one or more hidden layers, and an output layer. The nodes of each layer are connected to the nodes in the adjacent layers by weight parameters, which are determined by fitting FPMD results. The value y_j^l of node j in layer l is given by

$$y_j^l = f_j^l \left(B_j^l + \sum_i w_{l,j}^{k,l} \cdot y_i^k \right), \quad (1)$$

where $w_{l,j}^{k,l}$ is the weight parameter from node i in layer k to node j ($=i+1$) in layer l ($=k+1$), y_i^k is the input value from the previous layer k, and B_j^l is the bias weight, which is used as an adjustable parameter to shift the activation function f_j^l. The activation function used in this study is given by

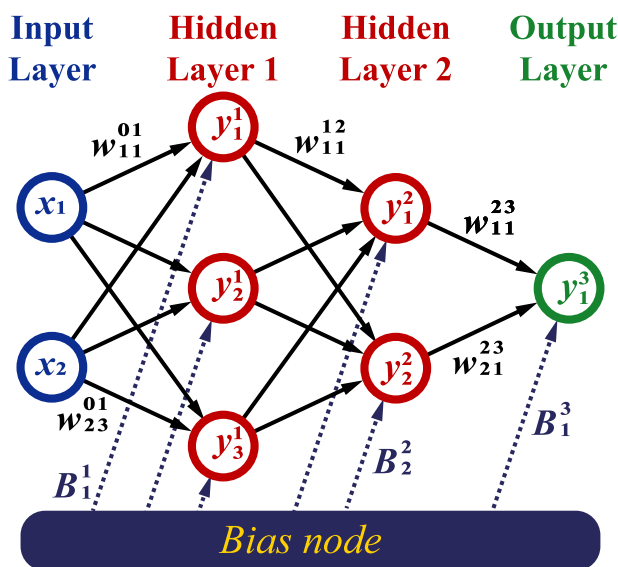$$f^l(x) = a \tanh(bx), \quad (2)$$

FIG. 1. Structure of a simple feedforward neural network. The value in the output layer (green) is obtained as a function of the input values x in the input layer (blue). There are several hidden layers comprising nodes (red). The black arrows represent the weight parameters $w_{l,j}^{k,l}$. The bias weights B_j^l (purple dashed arrows) are used as adjustable parameters to shift the activation function.

where $a = 1.7195$ and $b = 2/3$. This activation function is the modified hyperbolic tangent function to prevent the vanishing gradient during back propagation proposed by Lecan et al.[30]

In this study, we employ the FFNN comprising two hidden layers with 20 nodes for Ag and Se.

B. Symmetry functions

The total energy predicted by the ANN potential E^{ANN} is a sum of the atomic energies,[3]

$$E^{\text{ANN}} = \sum_i^{N_{\text{atom}}} E_i, \quad (3)$$

where E_i is the energy associated with the ith atom and N_{atom} is the total number of atoms in the system. Figure 2 shows a schematic of the ANN-potential construction. At first, the Cartesian coordinates $\{\mathbf{R}_i\}$, which are obtained by the FPMD simulation, are transformed into a set of symmetry function values $\mathbf{G}_i$ associated with each atom. These values depend on the positions of all atoms in the system. After the $\mathbf{G}_i$ values are used as the input values x for the FFNN (Fig. 1), the atomic energies E_i are obtained as the output values. The total energy E^{ANN} is then obtained by the sum of E_i, as shown in Eq. (3).

Symmetry functions were introduced by Behler[31] and have been used for the construction of various ANN potentials[32,33] and other machine-learning tasks.[34] There are two types of symmetry functions: a radial function G_i^{rad} and an angular function G_i^{ang}. The radial function G_i^{rad} is given by

$$G_i^{\text{rad}} = \sum_j e^{-\eta (R_{ij} - R_c)^2} \cdot f_c(R_{ij}), \quad (4)$$

where $R_{ij} = |\mathbf{R}_{ij}|$ is the distance between the ith and jth atoms, η and R_s are adjustable parameters, and $f_c(R_{ij})$ is the cutoff function defined as

$$f_c(R_{ij}) = \begin{cases} 0.5 \left[\cos \left(\frac{\pi R_{ij}}{R_c} \right) \right] & (R_{ij} \leq R_c), \\ 0 & (R_{ij} \geq R_c) \end{cases}, \quad (5)$$
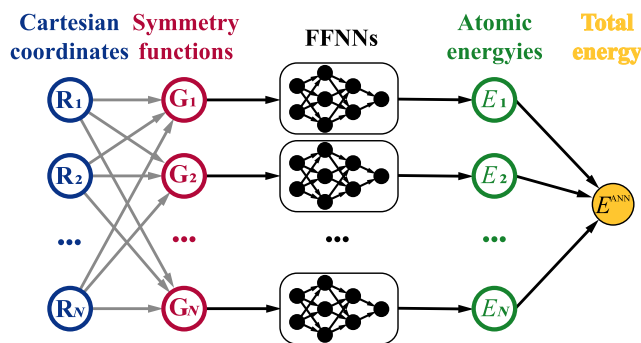

FIG. 2. Schematic of ANN potential construction.[3] The Cartesian coordinates $\{\mathbf{R}_i\}$ (blue) are transformed into a set of symmetry function values $\mathbf{G}_i$ (red) associated with each atom. These values depend on the position of all atoms in the system indicated by the gray arrows. The $\mathbf{G}_i$ values are used as the input values for the FFNNs, and the atomic energies E_i (green) are obtained as the output values. The total energy E^{ANN} (yellow) is obtained by summing E_i.

where R_c is the cutoff distance. The angular function G_i^{ang} is expressed as

$$G_i^{\text{ang}} = \sum_{j,k\neq i}^{\text{all}} \left(1 + \lambda\cos\theta_{ijk}\right)^{\zeta} \cdot e^{-\eta\left(R_{ij}^2 + R_{ik}^2 + R_{jk}^2\right)}$$
$$\cdot f_c(R_{ij}) \cdot f_c(R_{ik}) \cdot f_c(R_{jk}), \tag{6}$$

where θ_{ijk} is the angle between $\mathbf{R}_{ij}$ and $\mathbf{R}_{ik}$ with λ, ζ, and η being adjustable parameters.

The creation method of the symmetry functions used in this study and the list thereof are shown in Sec. I of the supplementary material. The parameters of the symmetry functions are listed in Tables S I and S II of the supplementary material. We processed the symmetry functions used in the previous study that created ANN potentials for TiO_2 crystals by Artrith and Urban[16] and defined 56 of those for Ag and Se.

C. FPMD simulation for preparing training data

FPMD simulations are carried out using a highly parallelized plane wave DFT program, QXMD,[17,18] which calculates the electronic states using the projector-augmented-wave (PAW) method.[35,36] The generalized gradient approximation (GGA) is used for the exchange-correlation energy.[37] The plane-wave cutoff energies are 20.0 and 200.0 Ry for the electronic pseudowave functions and pseudocharge density, respectively. The projector functions are generated for $4d$, $5s$, and $5p$ of Ag and $4s$, $4p$, and $4d$ of Se.

To prepare training data of α-Ag_2Se, the system with 256 Ag and 128 Se atoms in a cubic supercell is used under periodic boundary conditions. The number density of atom (0.0461 Å^{-3}) is taken from the experiment.[38] The corresponding size of the supercell is 20.271 Å. The NVT ensemble by the Nosé-Hoover thermostat[39] is employed for FPMD simulations. The temperature is set to be 500 K. The equations of motion are solved via an explicit reversible integrator[40] with a time step of Δt = 2.42 fs.

1000 FPMD steps in the equilibrium state that $g(r)$ converged were used for training of the ANN potential. As mentioned in Sec. I as one of the guidelines, this is because the convergence of the $g(r)$ was selected as the indicator for determining the amount of training data. We believe that this criterion is physically relevant, as the convergence of $g(r)$ guarantees that the local structure is well characterized. Section II of the supplementary material explains that $g(r)$ converged with 1000 steps of FPMD data. Ith step data include total energy (E_I^{FPMD}), atomic forces ($\{\mathbf{F}_{I,i}^{\text{FPMD}}\}$), pressures ($\{P_{I,i}^{\text{FPMD}}\}$), and Cartesian coordinates ($\{\mathbf{R}_i\}$). Movie 1 of the supplementary material shows the atomic dynamics during the 1000 steps, where the pink and yellow spheres represent Ag and Se atoms, respectively. For convenience, the dataset of total energies is hereafter referred to as D_E and that of atomic forces as D_F.

However, a mean square displacement (MSD) observed for the diffusion of Ag was not converged due to the insufficient sampling. We performed another 4000-step FPMD simulation to converge the MSD. There was thus no guarantee that the MSD yielded by the ANN potential trained with 1000-step FPMD data matched that of the FPMD simulation. The comparison result is described in Sec. III.

D. Physical quantities calculated by ANN-MD simulations for α-Ag_2Se

As stated in Sec. I, the current issue with the conventional empirical interatomic potential of α-Ag_2Se is that the structure of Ag is not consistent with those of FPMD simulation and experiment.[22] To our knowledge, since the experimental data of $g(r)$ at 500 K have not been reported yet, $g(r)$ of Ag–Ag, Ag–Se, and Se–Se calculated from our FPMD simulation are compared to those yielded by ANN-MD ones in Sec. III. The agreement means that this issue can be resolved.

In addition, since the diffusion of Ag must be described as well to confirm the reproduction of its superionic behavior, the MSD is calculated to compare the results of FPMD and ANN-MD simulations. We also compare the diffusion coefficient of Ag at 500 K[38,41] by calculating the slope of the MSD.

An experimental value of constant pressure specific heat (C_p) at 500 K has been reported.[24] By performing a long time scale ANN-MD simulation, the convergence value of C_p is evaluated.

The calculation of physical quantities above was performed in the equilibrium state of ANN-MD simulations with the NVT or NPT ensemble. In order to realize these ensembles, specifically, this study employed the Nosé-Hoover thermostat[39] and the Parrinello-Rahman-type barostat with an invariant variable cell shape method.[42] All the ANN-MD simulations conducted in this study started from an atomic structure in the equilibration stage of the FPMD one. The first 1000 ANN-MD steps were omitted as the equilibration stage, and the subsequent data were used to calculate the physical quantities above. Thus, note that the 1001th step is shifted to the first step. The equilibration process of ANN-MD simulations is described in Sec. III of the supplementary material.

E. Cost function and training termination conditions

We first defined a cost function $\mathcal{L}$ to train the ANN, which consists of not only the loss function of total energy (first term) but also those of atomic force (second term) and pressure (third term),

$$\mathcal{L} = \frac{p_E}{2}\frac{1}{N_I}\sum_I^{N_I}\left(E_I^{\text{FPMD}} - E_I^{\text{ANN}}\right)^2$$
$$+ \frac{p_F}{2}\frac{1}{N_I}\sum_I^{N_I}\frac{1}{3N_{\text{atom}}}\sum_i^{N_{\text{atom}}}\left(\mathbf{F}_{I,i}^{\text{ANN}} - \mathbf{F}_{I,i}^{\text{FPMD}}\right)^2$$
$$+ \frac{p_P}{2}\frac{1}{N_I}\sum_I^{N_I}\frac{1}{6}\sum_j^{6}\left(P_{I,j}^{\text{ANN}} - P_{I,j}^{\text{FPMD}}\right)^2, \tag{7}$$

where N_I is the number of FPMD training data. The factor 6 of the pressure term means that the number of independent degrees of freedom of the pressure tensor. $\{\mathbf{F}_{I,i}^{\text{ANN}}\}$ and $\{P_{I,j}^{\text{ANN}}\}$ is the atomic forces and pressures predicted by the ANN potential and is derived by the derivatives of Ith total energy E_I^{ANN} with respect to the Cartesian coordinates ($\{\mathbf{R}_i\}$) and the cell tensor $\mathbf{h}$, respectively. Section IV of the supplementary material describes mathematical forms of $\{\mathbf{F}_{I,i}^{\text{ANN}}\}$ and $\{P_{I,j}^{\text{ANN}}\}$. Since the three loss functions differ in dimension and size, p_E, p_F, and p_P are introduced as adjustment parameters. As the cost function of Aenet[16] consists of only the first term, we added the second and third terms. Note that Zhang et al. have already employed the same cost function,[15] and defined the

third term using the virial tensor, but we called the term pressure term. This is just because we wanted to emphasize the word "pressure," not "virial," though we implemented it on Aenet in the form of virial (see Sec. IV of the supplementary material).

However, this cost function is relatively useless since one needs to establish the adjustment method of the three parameters above. As we will explain the reason in Subsection III A, total energy and atomic force loss functions are used for training ANN potentials, but not the pressure loss function in this study. Instead, importantly, the error assessment for pressure is conducted while training.

Two cases are tried here: (1) training with only D_E as the previous studies and (2) training with both D_E and D_F (see Subsection II C for the meaning of these symbols). The training of the ANN is performed by the Limited memory Broyden–Fletcher–Goldfarb–Shanno for Bound constrained optimization (L-BFGS-B) method[43] implemented in Aenet. The number of epochs is set to be 2000 for all training cases. The details of the time spent on training and the volumes of training data are described in Sec. V of the supplementary material. Here, 95% and the remaining 5% of all data are randomly used for training and testing, respectively. Hereafter, these are called "Train" and "Test" datasets, respectively.

In Sec. III, the error evaluation of ANN potentials is not in the form of Eq. (7). Instead, dividing into total energy, atomic force, and pressure loss functions, the root mean square error (RMSE) calculated from each loss function and correlation coefficient (CC) between the reference values of FPMD data and predicted ones by an ANN potential is employed. In case (1) described above, the convergence of all the RMSE values is used as the training termination condition. In case (2), on the other hand, the additional condition that all CCs are 0.99 or more is set.

F. FPMD data for correction (D_{Cor})

As the reason is described in Sec. III, in fact, with only the training data prepared in Subsection II C, one can probably create an incomplete potential. In MD simulations with an ensemble such as NVT, there is a very low probability (but it may happen) that two atoms approach extremely close. Since the training data contain very few such extreme incidents, ANN potentials are created with imperfect short-range potential surfaces.

In order to resolve this issue, additional FPMD data containing many extremely short interatomic bonds are prepared with the atomic coordinates in which a side of the cubic supercell is reduced to 17.5 from 20.271 Å. These data are called D_{Cor} hereafter. The reason for 25% reduction in the length of the side is that a lot of target interatomic distances are available. The interatomic bond distances targeted here are those by furthermore 0.1 Å smaller than the shortest distances of Ag–Ag, Ag–Se, and Se–Se included in the training data (i.e., 2.4, 2.2, and 3.0 Å, respectively). However, it should be noted that this configuration is considered to be far from the equilibrium state, and, therefore, the potential accuracy would be worse if only the data obtained from this configuration are included in the training. This is because the respective potential surfaces created from the data above and the original training one may not be connected. We use two different optimization methods that for the atomic structure to get more of the data of interest and that for cell size to get closer to the original size. In the former case of
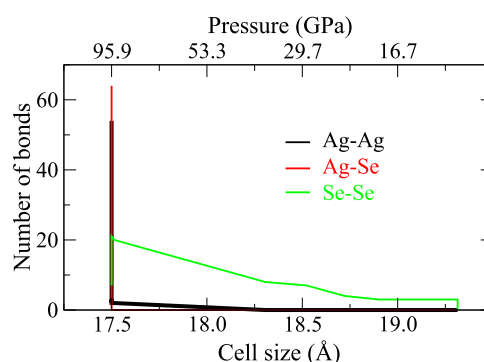

FIG. 3. The number of specific interatomic distances present in D_{Cor} as functions of cell size and pressure. The black, red, and green lines represent those between Ag–Ag, Ag–Se, and Se–Se, respectively, where the interatomic bonds within distances of 2.4, 2.2, and 3.0 Å were counted.

structural optimization (referred to as O$_{str}$), only the atomic coordinates are moved while the cell size remains fixed. At this time, since the distance between atoms does not greatly increase, it is possible to obtain target data of short interatomic distances. On the other hand, in the latter optimization of the cell size (referred to as O$_{cel}$), only the cell size is varied, and the relative atomic coordinates are thereby scaled. This is used to approximate the original cell size. Both optimizations were performed by the quasi-Newton method based on the BFGS algorithm implemented in our QXMD code.[17,18] We performed O$_{str}$ and O$_{cel}$ alternately 10 times, but the second O$_{str}$ was stopped by 5 times. The overall optimization profile was therefore O$_{str}$(10 times)–O$_{cel}$(10 times)–O$_{str}$(5 times). The total number of optimization steps was 25. The cell size resulted from 17.5 to 19.3 Å. The reason for having stopped optimization at 25 is that the interatomic distances between Ag–Ag, Ag–Se, and Se–Se present in the system became greater than 2.4, 2.2, and 3.0 Å, respectively. D_{Cor} was thus consisted of the 25 optimization-step data. The change in the number of Ag–Ag, Ag–Se, and Se–Se interatomic distances within 2.4, 2.2, and 3.0 Å, respectively, as functions of a side of the system and pressure in D_{Cor} is shown in Fig. 3. At least 50 interatomic distance data for each were available with this operation.

III. RESULTS

A. ANN potential without D_{Cor}

Here, we will describe the results of the trained ANN potential using the FPMD data of α-Ag$_2$Se at 500 K without D_{Cor}. We will illustrate that the ANN potential is incomplete but yields $g(r)$ and MSDs consistent with those of our FPMD simulation. In addition, the results of ANN potentials trained with only D_E and both D_E and D_F (referred to as $D_{E,F}$) are shown.

The profile of RMSEs and CCs of total energy, atomic force, and pressure during 2000 training epochs is shown in Fig. 4. Since all converged at 2000 epoch, Tables I and II show these RMSE and CC values at the epoch. In the conventional training with only D_E, the RMSE and CC of the total energy could be predicted extremely accurately. On the other hand, the CC of atomic force remained at about
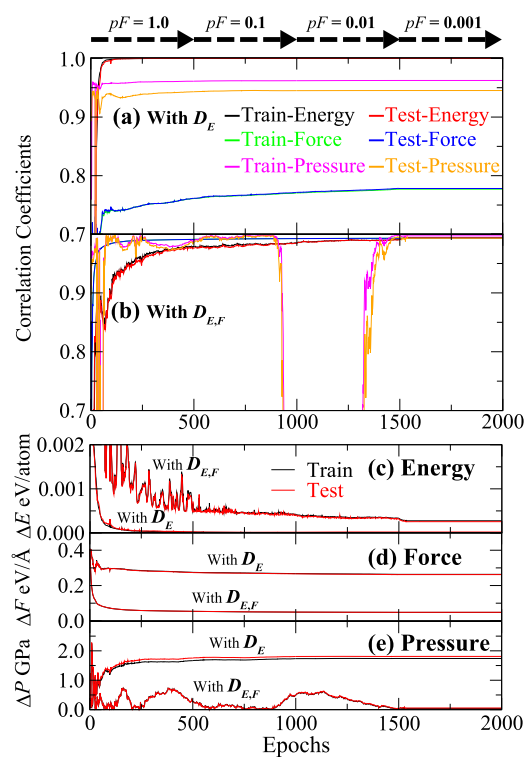
FIG. 4. The profiles of [(a) and (b)] CCs and [(c)–(e)] RMSEs of total energy, atomic force, and pressure during 2000-epoch training of ANN potentials with D_E and $D_{E,F}$ for Train and Test datasets. The top figure shows the change in the value of the coefficient of the loss function of atomic force p_F while training with $D_{E,F}$.

78% agreement for both Train and Test datasets, which was not a good prediction. The error of 1.8 GPa occurred in pressure even for the Train dataset as well. When an ANN-MD simulation under the NVT ensemble with the same supercell size as the FPMD one was performed, the structure could not be maintained from the beginning. One can check this in Movie 2 of the supplementary material. In addition, there are conserved quantities in energy unit in the MD

TABLE I. RMSEs and CCs of total energy (Energy), atomic force (Force), and pressure for Train and Test datasets at 2000 epoch of the ANN potential trained with D_E.

Root mean square error (RMSE)					
Energy (meV/atom)		Force (eV/Å)		Pressure (GPa)	
Train	Test	Train	Test	Train	Test
0.0060	0.0215	0.2634	0.2618	1.7392	1.8093
Correlation coefficient (CC)					
Energy (%)		Force (%)		Pressure (%)	
Train	Test	Train	Test	Train	Test
99.999	99.996	77.702	77.821	96.242	94.518

TABLE II. RMSEs and CCs of total energy (Energy), atomic force (Force), and pressure for Train and Test datasets at 2000 epoch of the ANN potential trained with $D_{E,F}$.

Root mean square error (RMSE)					
Energy (meV/atom)		Force (eV/Å)		Pressure (GPa)	
Train	Test	Train	Test	Train	Test
0.2754	0.2555	0.0480	0.0482	0.0420	0.0555
Correlation coefficient (CC)					
Energy (%)		Force (%)		Pressure (%)	
Train	Test	Train	Test	Train	Test
99.366	99.391	99.285	99.391	99.697	99.284

algorithms with explicit reversible integrators,[40] which are effective to check whether the dynamics of the system are physically relevant. The failure above was reflected in the conserved quantity of the ANN-MD simulation as a deviation [see around 1000 MD step (=2.42 ps) of the black line in Fig. 5(a)].

On the other hand, the ANN potential trained with $D_{E,F}$ showed good predictions, although the RMSE of the total energy decreased a little as shown in Fig. 4 and Table II. In addition, with respect to atomic force and pressure, the CCs achieved almost 1.0. In order to obtain this result, we adopted the method to gradually reduce the coefficient of the loss function of atomic force p_F in Eq. (7) during training while paying attention to the RMSE and CC of pressure. We first followed an example of the adjustment method of p_P as well as p_E and p_F proposed by Zhang et al.,[15] but it failed in our case. They might be a cause that the creating
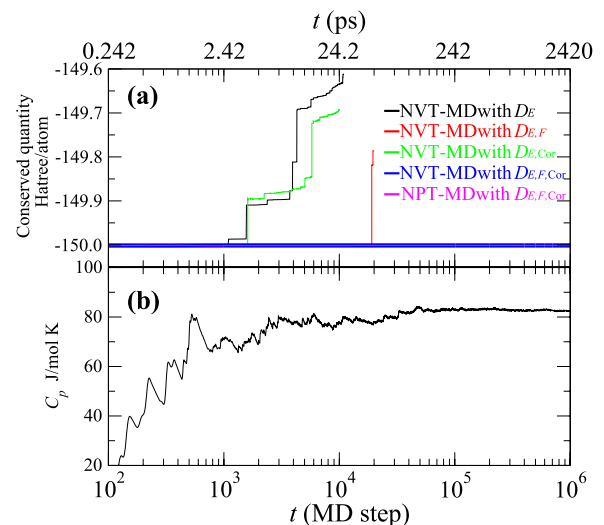

FIG. 5. (a) The time evolution of conserved quantities (Hartree/atom) and (b) constant pressure specific heat C_p (J/mol K) as a function of steps (or picoseconds) in our MD simulations using ANN potentials trained with D_E, $D_{E,F}$, $D_{E,\text{Cor}}$, and $D_{E,F,\text{Cor}}$. Here, MD simulations with NVT and NPT ensembles are referred to as NVT-MD and NPT-MD, respectively.

method of ANN such as the descriptor for the input layer and architecture differed from us. In addition, the adjustment method may depend on the system of interest, as the pressure is not correct unless both global (total energy) and local (atomic force) precisions are sufficient. The pressure loss function was therefore not used for training ANN potentials in this study (i.e., $p_P = 0$), resulting in a simpler training procedure by adjusting only p_E and p_F. [However, pressure errors (RMSE and CC) are constantly monitored during training.] Actually, we followed the training method with total energy and atomic force exemplified by Zhang *et al.*, which gradually reduced p_F during training.[15] We found that this method worked well, where, specifically, we set both coefficients p_E and p_F to 1.0 at first, and then reduced the p_F to 1/10 every 500 epochs (see the top of Fig. 4). This is understandable that the loss function of atomic force has much more freedom to be minimized than that of total energy when $(p_E, p_F) = (1.0, 1.0)$ so that the loss function of atomic force is mainly preferred as a target of training. Making p_F smaller means gradually shifting the training target to the total energy loss function. However, if this operation is performed without monitoring the error of pressure, as the CC around 1000 epoch shown in Fig. 4(b), one may not note that training is actually inadequate because total energy and atomic force show the good accuracies. At around 1000 epoch, it seems that the focus of training was still biased toward atomic force. By lowering p_F to eliminate the bias, this imbalanced issue was certainly resolved. Therefore, the pressure error must be always monitored, and this insight should be integrated to the adjustment method of Zhang *et al.* mentioned above.[15]

The ANN-MD simulation using this potential yielded the $g(r)$ and MSD as shown by red curves in Figs. 6 and 7, which were in good agreement with those of the FPMD one. It was also confirmed that the diffusion coefficient calculated from our MSD agreed with the experimental that of Ag at 500 K (1.5×10^{-5} cm^2/s).[38,41] Since there was no guarantee that MSD agreed with that of the FPMD
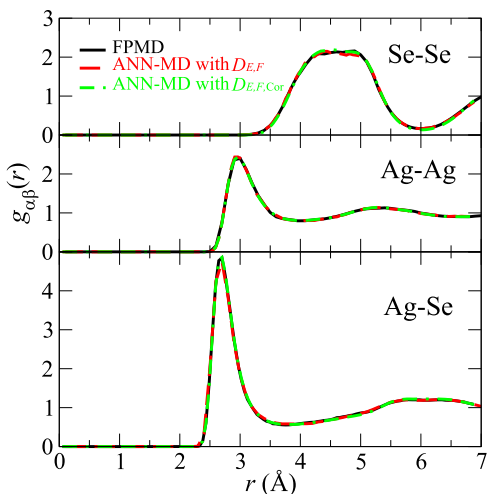

FIG. 6. The partial radial distribution functions $g_{\alpha\beta}(r)$ between Ag–Ag, Ag–Se, and Se–Se calculated by our FPMD and ANN-MD simulations using potentials trained with $D_{E,F}$ and $D_{E,F,Cor}$.
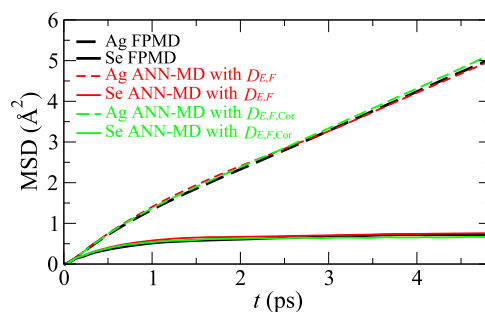

FIG. 7. The mean square displacements (MSD) (Å^2) of Ag and Se atoms calculated by our FPMD and ANN-MD simulations using potentials trained with $D_{E,F}$ and $D_{E,F,Cor}$. They were calculated on a time scale of 4000 times the MD time step Δt (=4.84 ps). The diffusion coefficients mentioned in the text were calculated from the slope after 3 ps.

simulation as stated at the end of Subsection II D, we consider that the convergence of $g(r)$ is a relevant criterion to decide the amount of training data. Movie 3 of the supplementary material shows the atomic dynamics during the beginning 1000 steps. Concerning calculation time, when converted to the time using the 1 CPU core, our FPMD simulation took 50 164.128 s/step, while the ANN-MD one took just 0.784 s/step. Therefore, we have achieved over 10^4 acceleration. This benchmark was performed with System B of the Institute of Physical Properties Research, the University of Tokyo, as noted in the acknowledgment.

However, there is still a problem with this ANN potential, and a failure occurred in the middle of the ANN-MD simulation. As shown by the red line in Fig. 5(a), a deviation of the conserved quantity can be seen after exceeding 19 000 steps (=46.0 ps). The $g(r)$ of Fig. 6 and the MSD of Fig. 7 were calculated by the data until 19 000 steps. In contrast to the catastrophic failure seen in MD with the ANN potential trained with D_E, there is, however, room for improvement on this potential. In the following subsection, we will describe the results of attempting to improve by adding correction data (D_{Cor}) to training data.

B. ANN potential with D_{Cor}

The cause of failure in the previous subsection was that a distance between any two atoms of Ag–Ag, Ag–Se, and Se–Se rapidly approaches and reached zero during ANN-MD simulations, resulting in the breakage of the whole atomic structure. The shortest distances for Ag–Ag, Ag–Se, and Se–Se included in the original FPMD training data were 2.5, 2.3, and 3.1 Å, respectively, and only a few have been identified. Since learning such rare events is quite difficult, it was assumed that the interaction between two atoms close to each other was not properly reproduced.

The training was thus performed by adding the correction data D_{Cor}, which is the FPMD data containing the number of interatomic distances 2.4, 2.2, and 3.0 Å for Ag–Ag, Ag–Se, and Se–Se, respectively. The values indicate the 0.1 Å shorter distances than the shortest ones in the original training data. The preparation of D_{Cor} was described in Subsection II F.

TABLE III. RMSEs and CCs of total energy (Energy), atomic force (Force), and pressure for Train and Test datasets at 2000 epoch of the ANN potential trained with $D_{E,\text{Cor}}$.

Root mean square error (RMSE)					
Energy (meV/atom)		Force (eV/Å)		Pressure (GPa)	
Train	Test	Train	Test	Train	Test
0.5187	1.0481	0.3651	0.5150	1.8314	3.5603
Correlation coefficient (CC)					
Energy (%)		Force (%)		Pressure (%)	
Train	Test	Train	Test	Train	Test
99.999	99.996	69.597	69.003	94.317	97.509

Here, we created ANN potentials by training with D_E and $D_{E,F}$ added D_{Cor} (referred to as $D_{E,\text{Cor}}$ and $D_{E,F,\text{Cor}}$) and then show the RMSEs and CCs at 2000 epoch in Tables III and IV. Even if we added D_{Cor}, prediction accuracies of atomic force and pressure were not improved for the case with $D_{E,\text{Cor}}$. When an ANN-MD simulation was performed with this potential, the atomic structure could not be maintained. The corresponding conserved quantity [the green line in Fig. 5(a)] has a deviation at around 1000 steps.

The ANN potential trained with $D_{E,F,\text{Cor}}$ was able to predict all properties reasonably, although the RMSE of pressure increased by one order compared with that of the potential trained with $D_{E,F}$ (see Tables II and IV). The reason for the reduced accuracy is that D_{Cor} contains the data that pressure reaches as high as about 100 GPa, as shown in Fig. 3. Conversely, this fact indicates that we could create an ANN potential that was able to predict such a wide range of pressures. Considering the appropriateness of the conserved quantity [see the blue line in Fig. 5(a)], the addition of D_{Cor} resolved the issue in the previous subsection and enabled an ANN-MD simulation on a million-step (=2.42 ns) scale. It was also confirmed that the $g(r)$ and MSD agreed with the results of our FPMD simulation, as shown in Figs. 6 and 7.

TABLE IV. RMSEs and CCs of total energy (Energy), atomic force (Force), and pressure for Train and Test datasets at 2000 epoch of the ANN potential trained with $D_{E,F,\text{Cor}}$.

Root mean square error (RMSE)					
Energy (meV/atom)		Force (eV/Å)		Pressure (GPa)	
Train	Test	Train	Test	Train	Test
0.5404	0.6029	0.0509	0.0556	0.2907	0.5404
Correlation coefficient (CC)					
Energy (%)		Force (%)		Pressure (%)	
Train	Test	Train	Test	Train	Test
99.998	99.998	99.231	99.433	99.942	99.958

While the $g(r)$ and MSD were converged with statistics of about several thousand MD steps, the specific heat that requires much more numerous steps is also calculated here. The constant pressure specific heat C_p under the ambient pressure at 500 K has been experimentally reported to be 83.6 J/mol K.[24] However, the ANN-MD simulations up to here were based on the NVT ensemble as well as our FPMD simulation, which provides only constant volume specific heat C_v. To calculate C_p, an ANN-MD simulation with the NPT ensemble was performed for a million steps. This also had the meaning of examination of whether MD simulations with the NPT ensemble can be performed using an ANN potential created from the FPMD data with the NVT ensemble. The profile of conserved quantity [magenta line in Fig. 5(a)] indicates that the robust simulation was performed. From the profile of specific heat shown in Fig. 5(b), we obtained the converged value to be C_p = 83.0 J/mol K, which is in excellent agreement with the experimental one.

We are also interested in how this ANN potential behaves under different conditions from the training data, in order to elucidate the atomic mechanism that generates anomalous increases in specific heat and thermal conductivity around the phase transition temperature of Ag_2Se (406 K). Therefore, in addition to 500 K, ANN-MD simulations under NPT-MD ensemble were performed at 450, 410, and 400 K. As our future expectations are included, the results were discussed in Sec. VI of the supplementary material, where the following three findings were described:

1. The specific heats agreed with the experimental values except that of 410 K. It was considered that the structural phase changes occurred even at 410 K because the system consisted of a small number of atoms (384 atoms).

2. At 400 (and 410 K), the system experienced several structural changes and resulted in a "β-like structure" (we will mention the meaning in the next finding). The key factor for the ANN-MD simulation being able to continue for 1×10^6 steps (=2.42 ns) without failures is that D_{Cor} would ensure the required physical property (i.e., interaction between two atoms that are close to each other).

3. The reason why we called the structure observed at 400 K "β-like structure" is that Ag atoms were diffused. Since β-Ag_2Se is not a superionic conductor, the observed one was thus not the complete β phase structure. However, this also means that the data at 500 K of the α phase contained information on the β structure even though it is not perfect. We expect that how much information is embedded in the data of 500 K can be quantified by uncertainty in the output of the ANN potential derived from the active learning technique, where the uncertainty is quantified using the variance of the values output from multiple ANN potentials with different initial weights that have learned the same training data.

Since the ANN is known to be vulnerable to extrapolation,[44] it would be difficult to reproduce physical phenomena that are not included in the training data. However, taking advantage of this weak point, such a template ANN potential trained under the specific condition may be useful, in order to infer what are the lacking data for reproducing physical phenomena under different conditions from the training data.

C. Training times and prospects for its mitigation

In this study, since the original Aenet[16] was extended (referred to as extended Aenet) as described in Subsection II E, the training times were longer than the original one. Finally, we thus describe how much training times have been increased and the prospects for its mitigation. Strictly, the training times and the volumes of training data are different from those of the original Aenet as described in Sec. V of the supplementary material, but only training times are mentioned here. Table V shows the training time using the original Aenet and those for four ANN potentials created in this study using our extended one. These computations were performed with System B of the Institute of Physical Properties Research, the University of Tokyo, and the number of CPU cores used was 96. The number of training epochs was set to be 2000. The training data that can be used in the original Aenet are only the total energy (i.e., D_E), and the corresponding training time was 3274 s. On the other hand, training with D_E using our extended Aenet took 106 793 s, about 33 times that of the original one. This time difference is due to the original Aenet output errors such as RMSE between the total energies predicted by the ANN potential (i.e., $\{E_I^{ANN}\}$) and those of FPMD at each epoch, while our extended Aenet does not only $\{E_I^{ANN}\}$ but also predicted atomic forces (i.e., $\{\mathbf{F}_{I,i}^{ANN}\}$), pressures (i.e., $\{P_{I,j}^{ANN}\}$), and their errors with those of FPMD (see Fig. 4), although the accuracy of ANN potentials created by both versions is completely consistent because the training data are the same. Whereas the computation of $\{E_I^{ANN}\}$ can be fast performed according to Eq. (3), those of $\{\mathbf{F}_{I,i}^{ANN}\}$ and $\{P_{I,j}^{ANN}\}$ take a quite long time due to the complexity of summations shown by Eqs. (IV.1) and (IV.2) in the supplementary material. However, as described in Subsection III A, if the errors of $\{\mathbf{F}_{I,i}^{ANN}\}$ and $\{P_{I,j}^{ANN}\}$ are not evaluated, one cannot figure out whether balanced training is achieved in two aspects, global and local. We thus must calculate the errors in exchange for an increase in training time. Actually, since these errors are not used for updating ANN weights, they can be calculated at a frequency specified by the user. Although we calculated every epoch in this study, for example, if every 10 epoch, the error calculation time 103 519 (=106 793 – 3274) s becomes its 1/10. Hence, in the case where only D_E is the training target, the training time can be easily reduced.

On the other hand, considering the balance of training, the case of using atomic force data (i.e., D_F) in addition to D_E (i.e., $D_{E,F}$) is more important. However, since $\{\mathbf{F}_{I,i}^{ANN}\}$ must be calculated for each epoch to update the weights of the ANN, it is not easy to reduce

the training time compared to the case of only D_E. The corresponding training time was 184 621 s. In addition, when the correction data D_{Cor} were added to $D_{E,F}$ (i.e., $D_{E,F,Cor}$), the training time further increased to 215 548 s (=about 2.5 days). D_{Cor} is difficult to train because this is the extreme dataset that contains atomic structures showing as high as 100 GPa (see Fig. 3). This is the longest training time in this study, and we consider that it is reasonable in terms of practicality for now if training is completed in about 2.5 days. However, in the future, it may be necessary to re-examine the training method when handling data that is several times the size of this study. The shortcut for solving this issue in terms of software is to reduce the number of training epochs by utilizing and improving optimization algorithms such as Levenberg-Marquardt, adaptive moment estimation, and Kalman filter methods, which are or are being implemented in the original Aenet. We will investigate optimization algorithms suitable for training using $D_{E,F,Cor}$ as future work.

IV. CONCLUSIONS

We have given an example for the creation of ANN potentials that can perform a long time scale MD simulation under at least the same conditions from FPMD data of specific conditions. The procedure is as follows: (1) enough FPMD data are prepared for $g(r)$ to converge; (2) while training the ANN with D_E and D_F, not only errors of total energy and atomic force but also those of pressure are monitored until the errors are minimized; and (3) more robust ANN potentials that do not break down in the middle of MD simulations can be created by mixing a small amount of FPMD data for expressing the interaction between atoms that are close to each other (D_{Cor}). As an application example, we created ANN potentials for α-Ag$_2$Se, and it was confirmed that the $g(r)$, MSD (or diffusion coefficient), and the specific heat requiring a long-time simulation reproduced the FPMD and the experimental values. In addition, the ANN-MD simulation achieved over 10^4 acceleration over the FPMD one. Therefore, we consider that this is an important first step to clarify the atomistic mechanism of the spike of specific heat and thermal conductivity accompanying the phase transition toward the low temperature β phase.[20,23,24] In addition, with some guidelines proposed in this study, a lot of FPMD data sleeping on hard disks after research may be utilized again. We believe that these findings can generally be used in the creation of the ANN potential for a lot of materials.

However, we still have to say that there are many untested issues regarding the applicable scope of these guidelines. The convergence condition of the $g(r)$ means at least that it is not assumed to utilize nonequilibrium FPMD data as training data. We attempt to set the definition of data preparation guidelines for such nonequilibrium states as one of the future works. Besides, for α-Ag$_2$Se, the transferability of our ANN potentials should be clarified by performing MD simulations under the conditions different from the training data. From the results of our trial (described at the end of Subsection III B), since D_{Cor} seems to prevent the catastrophic failures as shown in Fig. 5(a), it would provide accurate ANN potentials under the conditions near training data. On the other hand, the conditions far from the training data, such as β-Ag$_2$Se, may not be reproduced. However, using the ANN potential created in this study as a template, it is expected that what are essential data to

TABLE V. Training times for four ANN potentials created in this study using the extended version and for the one using the original Aenet.[16] "Dataset" indicates the dataset used to train the target ANN potential.

Version of Aenet	Dataset	Training time (s)
Extended	D_E	106 793
	$D_{E,F}$	184 621
	$D_{E,Cor}$	136 471
	$D_{E,F,Cor}$	215 548
Original	D_E	3 274

reproduce the β phase structure can be elucidated by the evaluation of the uncertainty defined in the active learning.

For systems with a high degree of freedom, only the data for expressing the interaction between two atoms approaching each other may be insufficient. For example, although rare, proton transfer occurs between water molecules even under an equilibrium state. As such, essential but rare incidents associated with the surroundings of molecules may cause failure of ANN-MD simulations, we would like to clarify what kinds of data should be prepared considering the previous studies for the water and aqueous electrolyte solutions.[2,6,7] If this becomes clear, ANN-MD simulations with FPMD precision for biochemical systems come into view.

SUPPLEMENTARY MATERIAL

See the supplementary material for detailed descriptions regarding the symmetry functions used in our study (Sec. I) and the convergence of the radial distribution function (Sec. II). The initial condition and the forms of atomic force and pressure tensor for the ANN-MD simulations are also described in Secs. III and IV, respectively. In addition, the training times and volumes of training data are summarized in Sec. V. Section VI describes the calculation result of specific heats at different temperatures from that of the training data and the characteristic atomic structures observed in the simulations.

REFERENCES

[1] J. Chen, G. Zhang, and B. Li, Phys. Lett. A **374**, 2392 (2010).
[2] J. Behler, Angew. Chem., Int. Ed. **56**, 12828 (2017).
[3] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[4] S. A. Ghasemi, A. Hofstetter, S. Saha, and S. Goedecker, Phys. Rev. B **92**, 045131 (2015).
[5] S. Faraji, S. A. Ghasemi, B. Parsaeifard, and S. Goedecker, Phys. Chem. Chem. Phys. **21**, 16270 (2019).
[6] T. Morawietz, A. Singraber, C. Dellago, and J. Behler, Proc. Natl. Acad. Sci. U. S. A. **113**, 8368 (2016).
[7] B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, Proc. Natl. Acad. Sci. U. S. A. **116**, 1110 (2019).
[8] M. Hellstrm and J. Behler, Phys. Chem. Chem. Phys. **19**, 82 (2017).
[9] V. Quaranta, M. Hellstroem, and J. Behler, J. Phys. Chem. Lett. **8**, 1476 (2017).
[10] T. S. Hy, S. Trivedi, H. Pan, B. M. Anderson, and R. Kondor, J. Chem. Phys. **148**, 241745 (2018).
[11] K. T. Schutt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K. R. Muller, J. Chem. Theory Comput. **15**, 448 (2019).
[12] E. V. Podryabinkin and A. V. Shapeev, Comput. Mater. Sci. **140**, 171 (2017).
[13] E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, and A. R. Oganov, Phys. Rev. B **99**, 064114 (2019).
[14] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E, Phys. Rev. Mater. **3**, 023804 (2019).
[15] L. Zhang, J. Han, H. Wang, R. Car, and W. E, Phys. Rev. Lett. **120**, 143001 (2018).
[16] N. Artrith and A. Urban, Comput. Mater. Sci. **114**, 135 (2016).
[17] F. Shimojo, R. K. Kalia, A. Nakano, and P. Vashishta, Comput. Phys. Commun. **140**, 303 (2001).
[18] F. Shimojo, S. Fukushima, H. Kumazoe, M. Misawa, S. Ohmura, P. Rajak, K. Shimamura, L. Bassman, S. Tiwari, R. K. Kalia, A. Nakano, and P. Vashishta, SoftwareX **10**, 100307 (2019).
[19] Y. Ding, Y. Qiu, K. Cai, Q. Yao, S. Chen, L. Chen, and J. He, Nat. Commun. **10**, 841 (2019).
[20] H. Chen, Z. Yue, D. Ren, H. Zeng, T. Wei, K. Zhao, R. Yang, P. Qiu, L. Chen, and X. Shi, Adv. Mater. **31**, 1806518 (2019).
[21] F. Drymiotis, T. W. Day, D. R. Brown, N. A. Heinz, and G. Jeffrey Snyder, Appl. Phys. Lett. **103**, 143906 (2013).
[22] F. Kirchhoff, J. M. Holender, and M. J. Gillan, Phys. Rev. B **54**, 190 (1996).
[23] S. A. Aliev and F. F. Aliev, Semiconductors **42**, 394 (2008).
[24] F. Grønvold, S. Stølen, and Y. Semenov, Thermochim. Acta **399**, 213 (2003).
[25] G. Cybenko, Math. Control Signals Syst. **2**, 303 (1989).
[26] K. Hornik, M. Stinchcombe, and H. White, Neural Networks **2**, 359 (1989).
[27] C. M. Handley and P. L. A. Popelier, J. Phys. Chem. A **114**, 3371 (2010).
[28] S. Lorenz, M. Scheffler, and A. Gross, Phys. Rev. B **73**, 115431 (2006).
[29] S. Lorenz, A. Gross, and M. Scheffler, Chem. Phys. Lett. **395**, 210 (2004).
[30] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural Networks: Tricks of the Trade*, 2nd ed., edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer Berlin Heidelberg, 2012), pp. 9–48.
[31] J. Behler, J. Chem. Phys. **134**, 074106 (2011).
[32] G. C. Sosso, G. Miceli, S. Caravati, J. Behler, and M. Bernasconi, Phys. Rev. B **85**, 174103 (2012).
[33] N. Artrith and A. M. Kolpak, Comput. Mater. Sci. **110**, 20 (2015).
[34] E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras, and A. J. Liu, Phys. Rev. Lett. **114**, 108001 (2015).
[35] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[36] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[37] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[38] H. Okazaki, J. Phys. Soc. Jpn. **23**, 355 (1967).
[39] S. Nosé, Mol. Phys. **52**, 255 (1984).
[40] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, Mol. Phys. **87**, 1117 (1996).
[41] M. Kobayashi, Solid State Ionics **39**, 121 (1990).
[42] G. J. Martyna, D. J. Tobias, and M. L. Klein, J. Chem. Phys. **101**, 4177 (1994).
[43] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, SIAM J. Sci. Comput. **16**, 1190 (1995).
[44] Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).